AMENDMENTS TO THE SPECIFICATION

D1 Kindly substitute the forty-six (46) page original Sequence Listing with the enclosed seventy-five (75) page amended Sequence Listing.

On page 5, make the following amendments to the paragraph spanning lines 13-17:

D2 Figure 7 shows the genomic organization and partial genomic sequence (nucleotides 2221-11570 of SEQ ID NO: 13) of the KLK-L2 gene. Intronic sequences are not shown except for the splice junctions. Introns are shown with lower case letters and exons with capital letters (SEQ ID NO: 96). The start and stop codons are encircled and the exon –intron junctions are boxed. The translated amino acids of the coding region (SEQ ID NO: 14) are shown underneath by a single letter abbreviation. The catalytic residues are inside triangles. Putative polyadenylation signal is underlined.

On page 38, make the following amendments to the paragraph spanning lines 5-18:

D3 The KLK-L2 gene, as presented in Figure 7, is formed of 5 coding exons and 4 intervening introns, spanning an area of 9,349 bp of genomic sequence on chromosome 19q13.3-q13.4. The lengths of the exons are 73, 262, 257,134, and 156 bp, respectively. The intron/exon splice sites (mGT....AGm) and their flanking sequences are closely related to the consensus splicing sites (-mGTAAGT ...CAGm-) (32). The presumptive protein coding region of the KLK-L2 gene is formed of 879 bp nucleotide sequence (SEQ ID NO: 96) encoding a deduced 293-amino acid polypeptide (SEQ ID NO: 14) with a predicted molecular weight of 32 KDa. There are two potential translation initiation codons (ATG) at positions 1 and 25 of the predicted first exon (numbers refer to SEQ. ID. NO.13 and GenBank Accession #AF135028). It is assumed that the first ATG will be the initiation codon, since : (1) the flanking sequence of that codon (GCGGCCATGG SEQ ID NO: 89) matches closely with the Kozak consensus sequence for initiation of

D3

translation (GCC A/G CCATGG SEQ ID NO: 90) (33) and is exactly the same as that of the homologous zyme gene. At this initiation codon, the putative signal sequence at the N-terminus is similar to other trypsin-like serine proteases (prostase and EMSP) (Figure 9). The cDNA ends with a 328 bp of 3' untranslated region containing a conserved poly adenylation signal (AATAAA) located 11 bp up-stream of the poly A tail (at a position exactly the same as that of the zyme poly A tail)(11).

On pages 38-39, make the following amendments to the paragraph spanning page 38, line 38 through page 39, line 8:

D4

The mature 227-amino acid sequence of the predicted protein was aligned against the GenBank database and the known kallikreins using the "BLASTP" and "BLAST 2 sequence" programs. KLK-L2 is found to have 54% amino acid sequence identity and 68% similarity with the enamel matrix serine proteinase 1 (EMSP1) gene, 50% identity with both trypsin like serine protease (TLSP) and neuropsin genes and 47%, 46%, and 42% identity with trypsinogen, zyme, and PSA genes, respectively. The multiple alignment study shows that the typical catalytic triad of serine proteases is conserved in the KLK-L2 gene (H^{108}, D^{153}, and S^{245}) and, as the case with all other kallikreins, a well conserved peptide motif is found around the amino acid residues of the catalytic triad [i.e., histidine (WLLTAA*H*C SEQ ID NO: 91), serine(GD*S*GGP SEQ ID NO: 92), and aspartate(*D*LMLI SEQ ID NO: 93)] (10, 11).

On page 42, make the following amendment to the paragraph spanning lines 23-32:

D5

A putative new gene, formed of three exons, was predicted by computer analysis of the genomic sequence. The predicted exons were subjected to sequence homology search against the human EST database (dbEST) and revealed an EST clone (GenBank accession # AA583908) which exhibited 99% homology with the putative gene. This EST was obtained, purified and sequenced and the sequence was aligned by BLAST software (37) against the genomic area that contains the putative gene. An additional

exon, downstream of the predicted structure, was identified. The 3' end of the gene was verified by: (a) The presence of the serine residue (S) of the catalytic triad in a well-conserved region. This highly conserved motif (GDSGGP SEQ ID NO: 92) always occurs at the beginning of the last exon in all known kallikreins. (b) The presence of a stop codon that is in frame with the predicted amino acid sequence. (c) The presence of a 19-poly A stretch at the end of the EST that was not found in the genomic sequence.

On page 50, make the following amendments to the paragraph spanning lines 13-27:

Alignment of the amino acid sequence of the KLK-L4 protein (long form) against the GenBank database and the known kallikreins, using the BLAST algorithm (37), indicated that KLK-L4 has 51% amino acid sequence identity with the TLSP and zyme genes, 49% identity with KLK-L2 and 47% and 45% identity with PSA and KLK2 genes, respectively. Multiple alignment study shows that the typical catalytic triad of serine proteases is conserved in the KLK-L4 gene (H^{108}, D^{153}, and S^{245}) and, as is the case with all other kallikreins, a well conserved peptide motif is found around the amino acid residues of the catalytic triad [i.e. histidine (WLLTAA**H**C SEQ ID NO: 91), serine (GDSGGP SEQ ID NO: 92), and aspartate (**D**LMLI SEQ ID NO: 93)] (Figure 27) (1, 11, 13, 35). In addition, several other residues were found to be fully or partially conserved among the human kallikrein gene family, as further shown in Figure 27. To predict the phylogenetic relatedness of the KLK-L4 gene with other serine proteases, the amino acid sequences of the kallikrein genes were aligned together using the "Clustal X" multiple alignment program and a distance matrix tree was predicted using the Neighbor-joining/UPGMA method (Figure 29). Phylogenetic analysis separated the classical kallikreins (KLK1, KLK2, and PSA) and grouped KLK-L4 with zyme, TLSP, KLK-L3, neuropsin, and NES1 genes, consistent with previously published studies (41) and indicating that this group of genes probably arose from a common ancestral gene by duplication.

On pages 58-59, make the following amendments to the paragraph spanning page 58, line 24 through page 59, line 1:

Although the protein encoded by the KLK-L5 gene is unique, it has a high degree of homology with the other kallikrein-like genes. The KLK-L5 protein (classical form) has 48% amino acid sequence identity and 57% overall similarity with neuropsin, 46% identity with the normal epithelial cell-specific 1 gene product (NES1) and 38% identity with both PSA and hK2 proteins. Multiple alignment shows that the typical catalytic triad of serine proteases is conserved in the KLK-L5 protein (H^{62}, D^{108}, and S^{200}) (Figures 33 and 36). In addition, a well-conserved peptide motif is found around the amino acid residues of the catalytic triad as is the case with other serine proteases [i.e., histidine (VLTAA***H***C SEQ ID NO: 94), serine (GD***S***GGP SEQ ID NO: 92), and aspartate (***D***LRLL SEQ ID NO: 95)] (11, 12) (Figure 36). Figure 36 also shows other amino acid residues that are completely conserved between kallikreins and kallikrein-like proteins. To predict the phylogenetic relatedness of the KLK-L5 protein with other serine proteases, the amino acid sequences of the kallikrein proteins were aligned together using the "Clustal X" multiple alignment program and a distance matrix tree was predicted using the Neighbor-joining/UPGMA and Protpars parsimony methods. Figure 37 shows the phylogenetic analysis which separated the classical kallikreins (hK1, hK2, and PSA) and clustered KLK-L5 with NES1 and neuropsin proteins in a separate group away from other serine proteases, consistent with previously published studies (27, 41) and indicating that this group of genes probably arose from a common ancestral gene, by gene duplication.